

Minireview

DNA microarray technology: the anticipated impact on the study of human disease

Javed Khan, Michael L. Bittner, Yidong Chen, Paul S. Meltzer, Jeffrey M. Trent *

Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

Received 30 November 1998; accepted 20 January 1999

Keywords: DNA microarray technology; Medical research; Data uniformity; Microarray

Contents

1. Introduction	M17
2. All chips are not created equal	M18
3. So many ESTs to choose from, so little room on each slide	M18
4. Technology, technology, technology	M20
4.1. Instrumentation	M21
4.2. Biochemistry	M23
4.3. Informatics	M26
5. Summary	M27
Acknowledgements	M27
References	M27

1. Introduction

This review focuses on the DNA microarray technology, its preliminary results, and the presumed enormous future impact the technique will have on the study of human disease. The development of this technology is gaining increasing importance as a di-

rect result of the Human Genome Project, and especially due to the recent focus on establishing an increasingly massive collection of genes (expressed sequence tags, ESTs) for human and model organisms. Our belief is that the continued development of the microarray technology will be invaluable to the study of human disease. However, it is a striking reality that this important research tool remains largely restricted to the few laboratories that have developed expertise in this area, and a growing number of commercial concerns. *Ultimately the real value*

* Corresponding author. Fax: +1 (301) 402-2040;
E-mail: jtrent@nhgri.nih.gov

of microarray technology will only be realized when this approach is available to the widest possible number of scientists.

2. All chips are not created equal

Miniaturized DNA microarrays (called DNA chips) have evolved as an important genome research tool. One major type of DNA chip contains high-density arrays of short (≤ 30 nucleotides long) oligonucleotides immobilized by photolithography to a solid surface [1–4]. This approach, pioneered by Affymetrix, allows as many as 300 000 oligonucleotides to be arrayed in a $< 2 \text{ cm}^2$ area. The resulting hybridization pattern is captured with the aid of a fluorescent scanning device and associated software that allows a merging of the hybridization pattern with the precise sequence of the arrayed DNA molecules. The applications of this approach are varied, but have included ‘re-sequencing’ of DNA for mutation detection [5] and more recently expression analysis [6]. It has the distinct advantage of high density, and the distinct disadvantage (until the human genome is sequenced in its entirety) of an absolute requirement for the arrayed sequence to be known. This report will not deal further with this commercially available technologic approach for generating or using microarrays. The reader is referred to URL 1 (see p. M27) for further information and potential applications.

The focus of this review will be on an alternative approach to comparatively analyze genome-wide patterns of mRNA expression. This approach, termed cDNA microarrays, uses cDNA clone inserts (instead of oligonucleotides), which are robotically printed onto a glass slide and subsequently hybridized to two different fluorescently labeled probes. For mRNA expression studies, the ultimate goal is to develop microarrays with every gene in a genome against which mRNA expression levels can be quantitatively assessed. At the current time this approach allows the deposition of as many as 25 000 cDNAs on a single microscope slide. While limited in terms of the number of targets per cm^2 compared to DNA chips, cDNA microarrays are significantly more flexible in the printing process and are not limited by the necessity to know the sequence of the clone being

arrayed. The technology for cDNA microarray hybridization is based on a method whereby the probes are pools of fluorescently labeled cDNAs synthesized after isolating mRNA from cells or tissues in two states that one wishes to compare. The differentially labeled probes are hybridized to the microarray slide; the resulting fluorescent intensities are measured using a laser confocal fluorescent microscope; and ratio information is obtained following image processing.

Finally, database development and design are critical, in order to deal with the immense amount of information generated by cDNA microarrays. This minireview briefly considers aspects of clone-set identification, printing and reading instrumentation, standard protocols, image analysis, and the database management behind this technology. As an illustration, it will highlight the approach developed in our laboratory. The reader is referred to URL 2 for further information and specific details of the NIH Microarray Project, and an associated Web Links page URL 3 referencing other academic and corporate sites involved in this technology.

3. So many ESTs to choose from, so little room on each slide

Estimates of the total number of genes in the human genome range from 70 000 to $> 100\,000$ [7–10]. With the sequence of the human genome about 6% completed (F. Collins, personal communication) and with the completion of the first sequence of the human now anticipated by 2003, we should, in the next decade, identify most (if not all) of the genes of the human genome and other targeted model organisms (see URL 4) [11]. However, it is obvious that in order for an organism to function, each gene must be expressed in a specific temporal and spatial context. Genome-wide expression studies using microarray-based technologies are an important link between sequence and function. In addition to the large body of sequence information from genomic sequences, a large volume of EST sequence data (from many organisms) has been archived over the past decade. Now, as a result of this deposition of EST sequence information, there are more DNA sequences in GenBank than there are publications in the biomedical literature [12] (Fig. 1, URL 5).

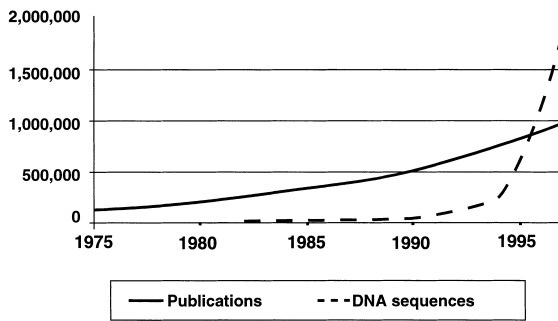


Fig. 1. Cumulative growth molecular biology and genetics literature compared with DNA sequences. Articles in the 'G5' (molecular biology and genetics) subset of MEDLINE were plotted along side of DNA sequence records in GenBank over the same time period. Used with permission [12].

The slope of this curve should continue to rise with future publications of expression profiles of thousands of genes in a single report (in contrast to the current practice of reporting the expression profile of a single gene).

Although sequence-based information on ESTs is

essential for genome-wide expression analyses, it is also problematic. Specifically, based on the posting of 8 October 1998 of the UniGene [13,14] (URL 6) collection of human expressed sequences, the total number of EST sequences numbers 797 691 (URL 7). With an estimated 100 000 human genes, and the certainty that all genes do not yet reside in these databases, it is obvious that significant redundancy exists in current EST databases. For practical concerns of economics and time, it is essential to identify a non-redundant collection of human and model system genes and ESTs. The UniGene collection of the National Center for Biotechnology Information of the NIH has currently identified such a set for human ESTs, with 40 000 genes (termed clusters) currently represented. A subset of 15 000 transcribed human sequences, which we refer to as the 15K set, has been compiled, and information about this set is available at URL 8. A similar UniGene set for murine ESTs has been generated, and other model organisms will follow. It is this non-redundant 'backbone' of ESTs that our laboratory (and many others) have used in selecting clones for arraying.

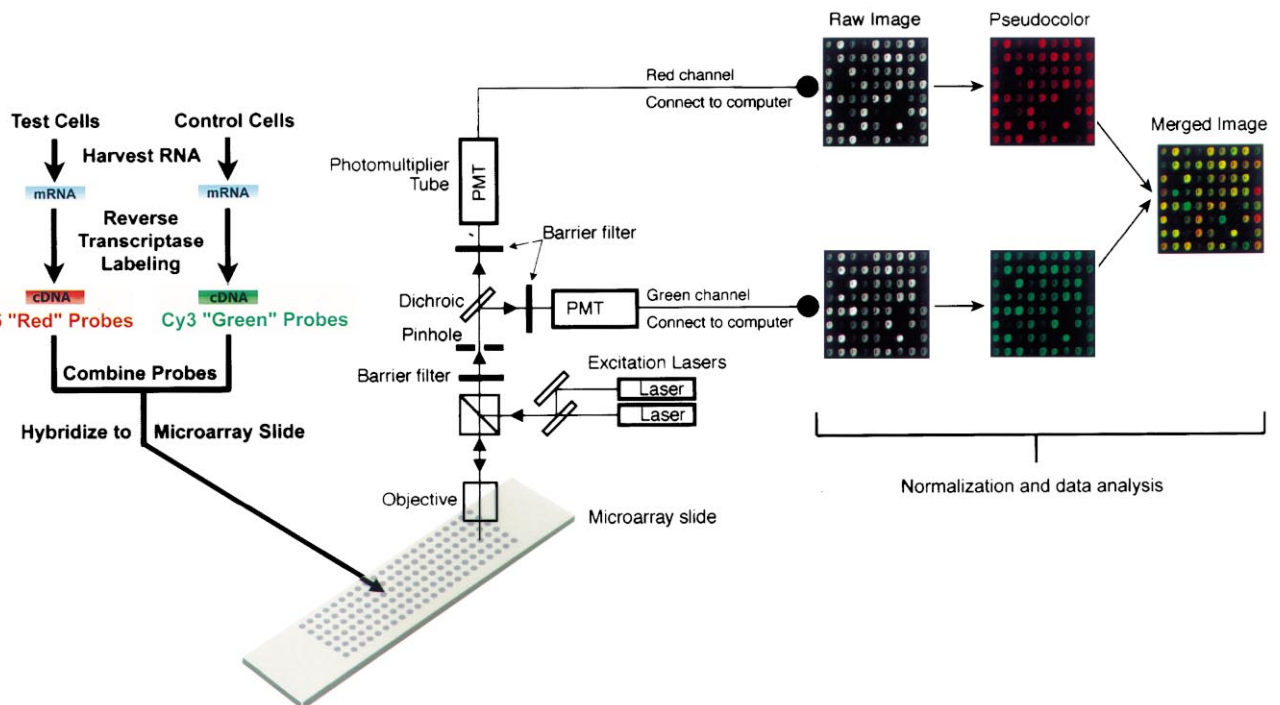


Fig. 2. Schematic of probe preparation, hybridization, scanning and image analysis.

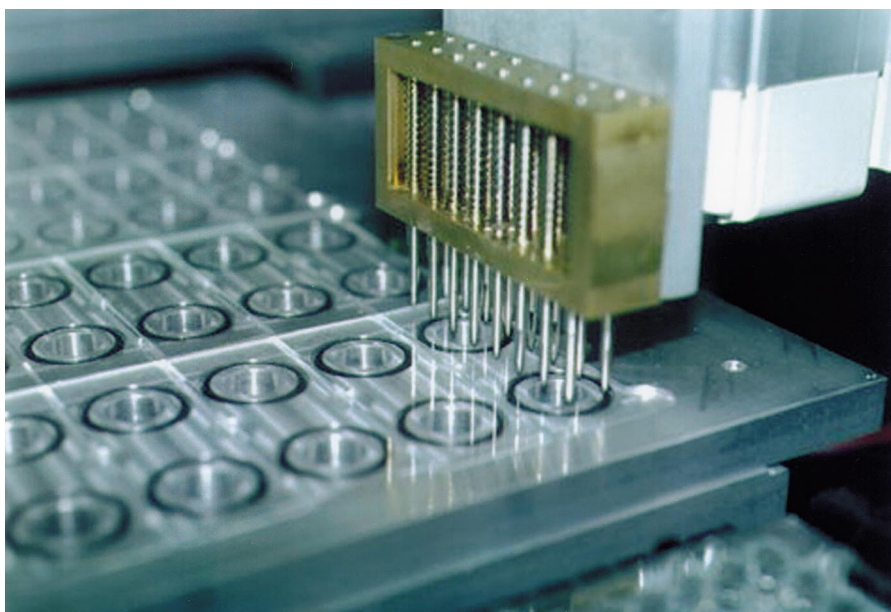


Fig. 3. Close view of print head loaded with 16 'quills'.

4. Technology, technology, technology

Regardless of the clone set chosen for any microarray, experiments are performed in much the same way. For the purposes of this minireview, we will

limit our comments to the cDNA microarray technology utilized in our laboratory at the National Human Genome Research Institute, NIH. Briefly, RNA is extracted from the test and control cells/tissue, purified, and then labeled by the reverse tran-

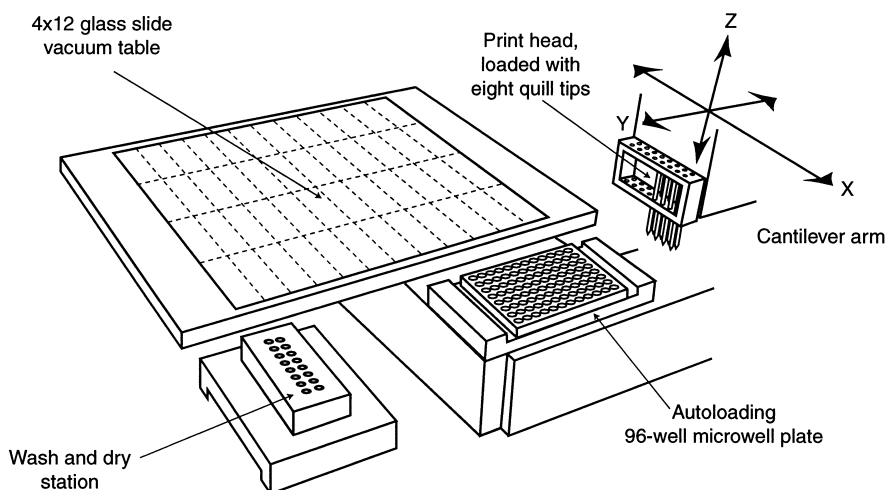


Fig. 4. cDNA microarray target printing apparatus. Computer-controlled robotic cantilever arm, capable of moving in XYZ directional planes, can be armed with up to 16 (two rows of eight) 'quill' print tips on the print head. In one automated print cycle, the print head dips the quills into a set of target DNA wells arrayed in 96 well microwell plates; then the print head traverses the vacuum table and touches the quill tips to each glass slide in succession, depositing target DNA; the print head continues to the wash/dry station where the tips are cleaned twice with water and dried. This cycle repeats as the print head returns to wet the tips in the next set of targets, continuing until all targets of a 96 well microwell plate have been printed. An autoloading mechanism removes spent microwell plates and can serve up new plates. By this method, microarray slides can be printed with as many as 15 000 precise and discrete cDNA targets. Used with permission [16].



Fig. 5. Entire arrayer showing the casing and controlling PC.

scriptase reaction to produce either fluorescently ‘tagged’ cDNA (Fig. 2). The labeled cDNA probes are then hybridized onto the array. Following hybridization, the array is scanned using a confocal laser scanning microscope, and the resulting data are analyzed.

The major issues affecting this technology can be divided into those related to instrumentation, biochemistry, and bioinformatics. Each will be considered briefly.

4.1. Instrumentation

Instrumentation remains a major and limiting factor at the present time, in that the most advanced expression systems are very expensive. As a result, they have only been available to laboratories in which this technology is being developed or to large commercial companies. Even filter-based radioactive systems, the lower cost alternative to fluorescent cDNA microarrays or DNA chips, are certainly not inexpensive to produce or to buy. To make all but the smallest filter array, one would need to construct a robot, order and amplify the clone set of cDNAs to array, re-sequence these to be sure of their

identity, and develop the technical approach to array them on membranes.

This is impractical for single investigators at most research centers, leaving the alternative, to pay an average of \$0.25 per clone for arrayed clone sets on nylon filters. Although these nylon filter arrays can usually be stripped and re-probed, the cost per experiment for large sets of ESTs can be prohibitive. Nevertheless quantitative expression data can be obtained by scanning these radioactive nylon arrays with a phosphorimager and analyzing the image data with relevant software (e.g. deArray, developed by Dr. Yidong Chen [15]) and other commercially available systems. This eliminates the need to buy an expensive fluorescence detector, and, as described below, lowers the amount of mRNA required for probe labeling. Until the cost per experiment falls, the number of experiments that can be performed and the number of replicates per experiment will be limited. In contrast to the printing of large arrays containing thousands of ESTs, it may be practical for groups of investigators to adapt the more widely available standard laboratory robots to the printing of a few hundred selected cDNAs (which can be obtained from public clone archives) that correspond

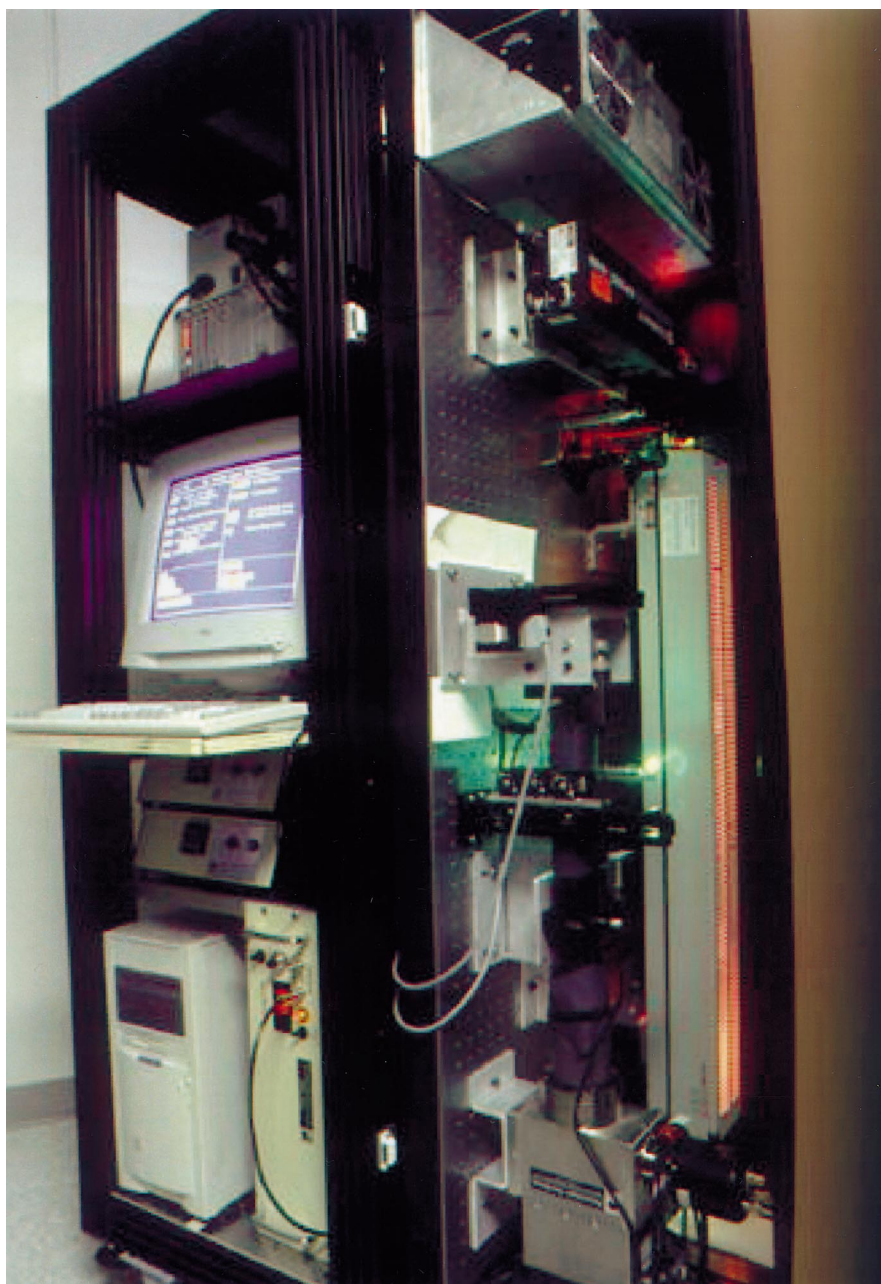


Fig. 6. Overview of confocal laser scanning microscope reader.

to genes involved in pathways of interest for particular projects.

Our laboratory at the NHGRI, in collaboration with the Biomedical Engineering and Instrumentation Program (BEIP) and the Division of Computer Research Technology (DCRT) at NIH, has developed the robotic instrumentation for arraying cDNA clones (termed arrayer), as well as a confocal

scanning fluorescent microscope to examine fluorescent intensities of microarrays (termed reader). As in the case of experimental protocols, the instrumentation we are using is under constant development in a research setting. Accordingly, the information provided merely illustrates the common components of arrayers and readers.

The arraying instrument was conceptually modeled

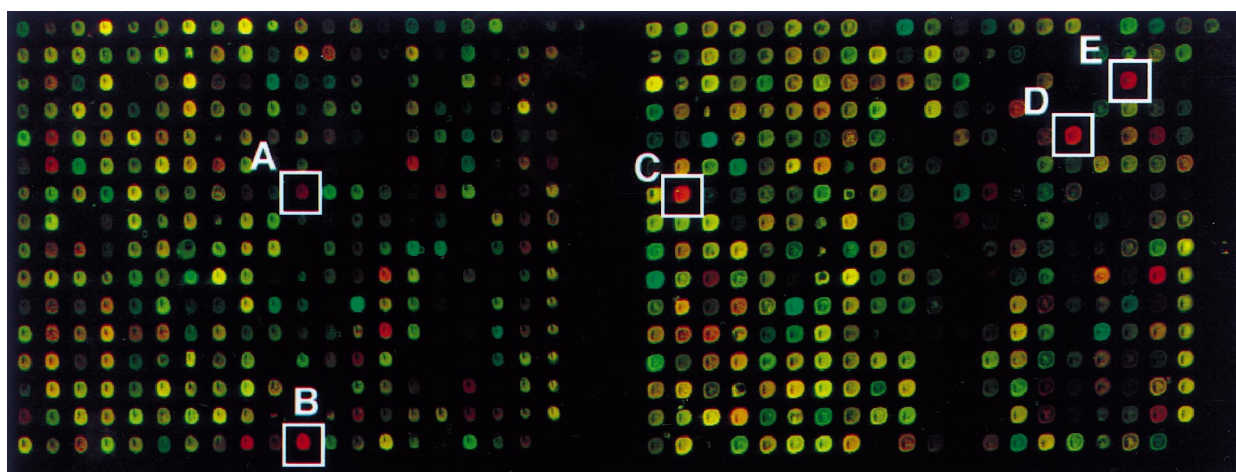


Fig. 7. Representative microarray hybridization. This pseudocolored image represents a portion of a microarray with the reference probe (normal fibroblasts) in green and rhabdomyosarcoma in red. The up (red) and down (green) regulation of several genes are illustrated. Representative genes of interest are boxed (A = FKHR, B = MYCN, C = CDK4, D = MYBL2 and E = NGFR). Ratio data for each of these individual spots are calculated and used for further analysis. Used with permission [19].

after that of Brown and colleagues (URL 9), and comprises an XYZ cantilever type robot holding 4–16 quill pens (Fig. 3, URL 10). Other features of the arrayer include the provision of a vacuum chuck for holding 48 standard microscope slides, a microtiter-tray loader/stacker, a wash/dry station (Fig. 4) [16], a controlling PC, air-handling components and a cabinet (Fig. 5, URL 11). The arrayer is able to put down in one cycle up to 16 spots on each of 48 slides, and wash and dry the quill pens for the next set of cDNAs, all in about 70 s. Most of this time is taken up with the actual spotting, since the wash and dry cycles take about 2 s each. Loading takes about 10 s. Thus, the contents of one 96 well tray can be spotted on 48 slides every 7 min, and 10 000 spots would take about 12 h.

The reader designed at the NIH is also based on the original design by Brown and colleagues (URL 9), although a number of additional features have been added. The reader is basically a computer-controlled inverted scanning fluorescent confocal microscope with a triple laser illumination system (Fig. 6, URL 12). The optical system is folded and arranged on an optical breadboard. The breadboard is hung with shock mounts in a vertical plane (to save space) inside a lightweight enclosure which also protects the optics from laboratory dust and personnel from laser light. Illumination is from three air-cooled lasers: a

488 nm, 100 mW argon ion laser for exciting FITC; a 532 nm, 100 mW NdYag for Cy3, and a 633 nm, 35 mW HeNe for Cy5. Any two lasers may be turned on simultaneously and their beams combined with dichroic mirrors and delivered to the specimen via a single dichroic and an objective lens (0.75 NA, 0.66 mm wd). The objective lens can be precisely focused with a digital controller. The emitted light from the fluorescent targets, after passing back through the objective and primary dichroic, is focused through a confocal pinhole and through a secondary dichroic onto two cooled PMTs which operate in parallel for the two different wavelengths. Data are acquired with a custom integrator and standard 16 bit A/D card in a PC. The operator can set the gain, speed, pixel size, pattern position, and pattern area. All of the electronics and power supplies are mounted in the cabinet bay next to the optics.

At 100 mm/s, with 20 μm^2 pixels, a 50 \times 20 spot array with spots on 400 μm centers involves 400 traverses each about 20 mm long and can be scanned in about 4 min. We can reliably detect about 10 pg/ μl of each species of cDNA [17].

4.2. Biochemistry

In an ideal world, to facilitate interpretation of

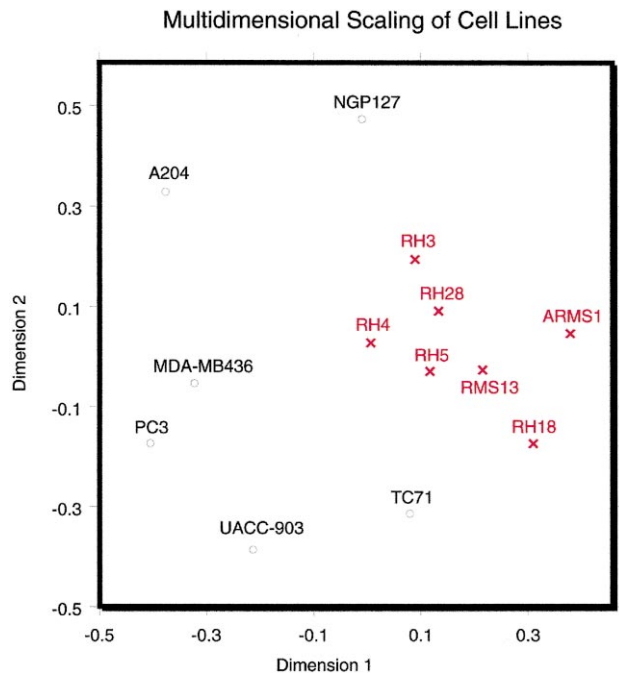


Fig. 8. The positions of all 13 cell lines in two-dimensional Euclidean distances are determined using the method of multidimensional scaling to make the between cell lines correspond as closely as possible to 1 minus the Pearson correlation coefficient of the log-ratio values. The *X* and *Y* scales are arbitrary. Cell lines falling close to one another in the plot have high correlation values. Using this method the ARMS cell lines (red) cluster together and the non-ARMS tumors (black) fall at the periphery of the plot. Used with permission [19].

expression data, the source of the mRNA to be studied should be isolated from a homogeneous population of cells, collected in a way that accurately preserves mRNA expression levels, and is available in an unlimited quantity. These criteria are best fit by the study of cultured cell lines, which can be derived from clonal populations and can be readily manipulated by pharmacological or physiological means for expression studies. In reality, one would like to be able to study human or animal tissues in addition to cell lines. However, tissues are difficult to study, because most are composed of numerous cell types, and the techniques and timing of sample collection vary, as does the quantity and quality of mRNA.

One way around the important issue of multiple cellular populations within a single biopsy specimen is to use the method of laser capture micro-dissection [18] to isolate specific cells for analysis. The limiting factor in this analysis is that very little RNA can be

prepared from the numbers of cells (<1000) obtained. Consequently, methods are being devised that permit reverse transcription, linear amplification, and alternative direct labeling strategies of vanishingly small amounts of RNA, which hopefully will maintain the relative concentrations of RNA species in the starting material.

Perhaps the most important advantage of fluorescent probes is the ability to use multiple spectrally distinct fluorophores. Because they can be multiplexed, hybridization signals from two or more differently labeled probes can be detected separately on the same slide. Thus, two-color fluorescence hybridizations allow direct comparison between two probes that were hybridized simultaneously to the same array. Ratio measurements also increase the accuracy of the comparative analysis. Finally, a series of protocols from our laboratory, which detail current protocols for mRNA isolation and probe labeling, can be found and downloaded at URL 13.

Hierarchical Clustering Dendrogram

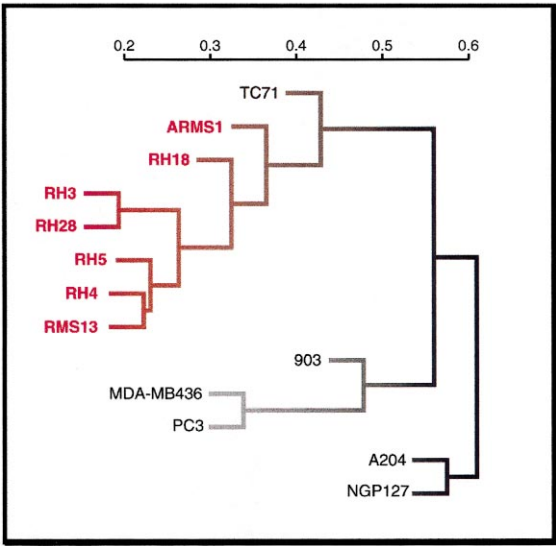


Fig. 9. The hierarchical clustering dendrogram indicates the order in which the 13 cell lines are combined to form clusters. The calculation of the dendrogram uses 1 minus Pearson correlation coefficient of log-ratios as the dissimilarity measure. The scale represents the distance between merged clusters and cell lines most similar are combined first. Using this method the seven ARMS cell lines (red) again cluster together. Used with permission [19].

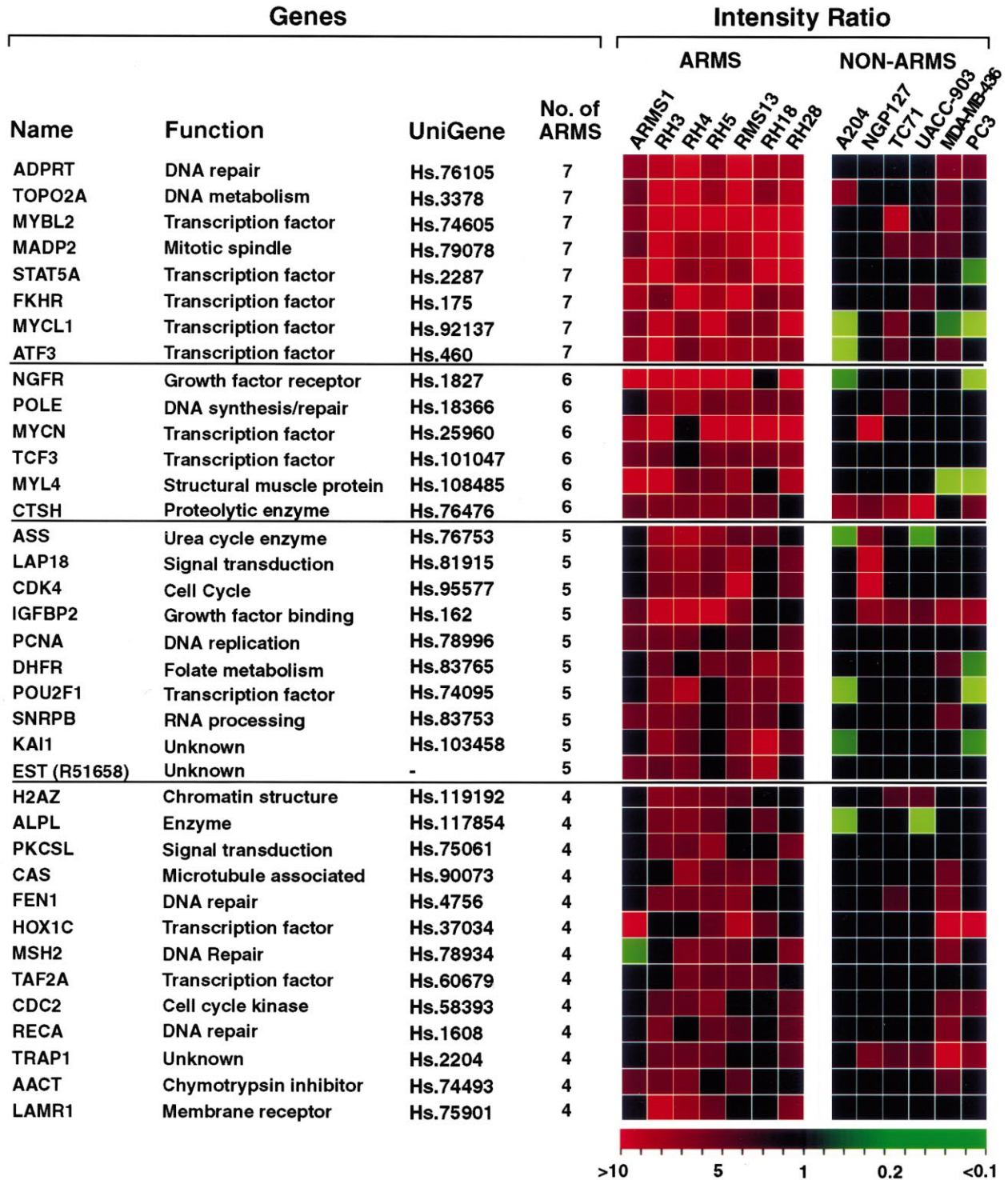


Fig. 10. Genes overexpressed in ARMS. Thirty-seven genes highly expressed relative to the reference probe in at least four out of seven ARMS and their functions are listed in the first and second columns. The third column indicates the number of ARMS cell lines in which each gene is up-regulated compared with the control at a level greater than the 99% confidence interval, and the fourth column provides the UniGene cluster designation. The expression ratios from each cell line are color-coded such that a red color indicates overexpression and green color reduced expression in the tumor compared with the control cell line. When the ratios of expression exceed the 99% confidence interval, the saturation of the red or green color increases in proportion to the ratio. A ratio color scale is shown at the bottom of the figure. Used with permission [19].

4.3. Informatics

It is obvious that large-scale, high-throughput experimental methods require information processing coupled to a variety of analysis tools. Laboratory Information Management (LIM) software and database systems to design arrays, to track clones, to collect, analyze, and interpret data from gene expression studies, are in their infancy. Among other things, such systems have to catalog the expression behavior of thousands of genes in a single experiment, and subsequently make comparisons across tissues, developmental and pathological states, or cellular perturbations. Very large quantities of data have to be managed both prior to and after an experiment, because direct access is required to all sequences, annotations, and physical DNA resources for the genes of the organism studied. Prior to analysis of data, the readout of relative expression levels observed on an array must be stored and preserved so that it is available for image processing, statistical, and, finally, biological analysis. The latter includes identifying transcripts that show statistically significant changes in absolute or relative levels of expression. Our laboratory has developed an image analysis program called deArray, developed by Dr. Yidong Chen [15], which handles, in seconds, the image capture and array spot identification steps, and catalogues the information on all fluorescent signals within a hybridization. The criteria for statistical significance and methods can be found at URL 14.

Very recently our laboratory has also published an array database management system, called ArrayDB, which provides a template for handling the flow of array data and presents several analysis components for ‘mining’ data generated by this technology. URL 15 depicts the elements of this system. The interested reader is referred to Ermolaeva and colleagues [12] for a text-based description and can download the schemata and software for this Sybase-anchored database at URL 13. This database performs the most obvious and straightforward of requirements including the provision of information (by hyperlinks to Entrez, NCBI) about any available sequence, structures, and functions of the gene products of interest. While these tools are both necessary and extraordinarily time-saving to the biologist, interpreting this

wealth of information clearly remains the responsibility of the investigator, who must be able to interrogate the data sets in multiple ways.

One approach we are incorporating within ArrayDB is a link to all biochemical pathways to which a particular transcript belongs and to genes with which the transcript is thought to interact. In the long run, our laboratory (and many others) are looking at ways to modify the software so that it is capable of pre-interpreting data (using a biochemical knowledge base and set of heuristics), presenting an investigator with alternative hypotheses or explanations of its meaning, and predicting relationships and possible pathways of interactions of the genes under study. It should be clear that there is an absolute requirement for computational biology, mathematical modeling, and developmental biostatistical approaches for successful biologic interpretation of the results of array experiments. This conjoining of disciplines is the only way that experiments involving hundreds of thousands of genes, with tens of thousands of changes across multiple experiments, can be managed.

Very recently, we utilized cDNA microarrays to investigate the gene expression profile of a group of seven alveolar rhabdomyosarcoma (ARMS) cell lines characterized by the presence of the *PAX3-FKHR* fusion gene and six unrelated controls [19]. This study illustrates the power and approaches of modeling data from array experiments. The study is based on the fact that several forms of human sarcoma, lymphoma, and leukemia are characterized by somatically acquired chromosome translocations which result in fusion genes encoding chimeric transcription factors with oncogenic properties. Fig. 7 shows an example of a microarray experiment where the expression profile of a normal myo-fibroblast cell line (green) was compared with that of an ARMS cell line (red). Using the method of multidimensional scaling (Fig. 8) to represent the relationships among the cell lines in two-dimensional Euclidean space, we found that ARMS cells show a consistent pattern of gene expression allowing them to be clustered together (Fig. 9). By searching across the seven ARMS cell lines, we found that a minimal subset of only 37 genes were most consistently expressed in ARMS relative to a reference cell line (Fig. 10). These results demonstrate the potential of cDNA microarray tech-

nology to elucidate a tumor-specific gene expression profile or ‘fingerprint’ in human cancers.

5. Summary

One can imagine that, one day, there will be a general requirement that relevant array data be deposited, at the time of publication of manuscripts in which they are described, into a single site made available for the storage and analysis of array data (modeled after the GenBank submission requirements for DNA sequence information). With this system in place, one can anticipate a time when data from thousands of gene expression experiments will be available for meta-analysis, which has the potential to balance out artifacts from many individual studies, thus leading to more robust results and subtle conclusions. This will require that data adhere to some type of uniform structure and format that would ideally be independent of the particular expression technology used to generate it. The pros and cons of various publication modalities for these large electronic data sets have been discussed elsewhere [12], but, practical difficulties aside, general depositing must occur for this technology to reach the broadest range of investigators.

Finally, as mentioned at the beginning of this review, it is unfortunate that this important research tool remains largely restricted to a few laboratories that have developed expertise in this area and to a growing number of commercial interests. Ultimately the real value of microarray technology will only be realized when this approach is generally available. It is hoped that issues including platforms, instrumentation, clone availability, and patents [20] will be resolved shortly, making this technology accessible to the broadest range of scientists at the earliest possible moment.

URLs.

- URL 1: <http://www.affymetrix.com/>
- URL 2: <http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/>
- URL 3: <http://www.nhgri.nih.gov/DIR/LCG/15K/links.html>
- URL 4: <http://www.nhgri.nih.gov/98plan/>
- URL 5: <http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/fig1.gif>
- URL 6: <http://www.ncbi.nlm.nih.gov/UniGene/>
- URL 7: <http://www.ncbi.nlm.nih.gov/UniGene/Hs.stats.shtml>

- URL 8: <http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/p15kabout.html>
- URL 9: <http://www.stanford.edu/pbrown/array.html>
- URL 10: http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/images/pins_jpeg.jpg
- URL 11: http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/images/arrayer_jpeg.jpg
- URL 12: http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/images/reader_jpeg.jpg
- URL 13: <http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/protocol.html>
- URL 14: http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img_analysis.html
- URL 15: http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/ng_paper.html

Acknowledgements

The custom-built robotic arrayer was developed by Stephen B. Leighton and scanner optics by Paul D. Smith. Thomas Pohida developed the electronics of both the arrayer and scanner. We thank Yuan Jiang, Gerald C. Gooden, John Lueders, Kim A. Gayton, Art A. Glatfelter and Robert L. Walker for their excellent technical assistance on this work. We also thank Lao H. Saal and Darryl Leja for assistance in the preparation of the illustrations.

References

- [1] R.J. Lipshutz et al., Using oligonucleotide probe arrays to access genetic diversity, *BioTechniques* 19 (1995) 442–447.
- [2] S.P. Fodor et al., Multiplexed biochemical assays with biological chips, *Nature* 364 (1993) 555–556.
- [3] S.P. Fodor et al., Light-directed, spatially addressable parallel chemical synthesis, *Science* 251 (1991) 767–773.
- [4] A.C. Pease et al., Light-generated oligonucleotide arrays for rapid DNA sequence analysis, *Proc. Natl. Acad. Sci. USA* 91 (1994) 5022–5026.
- [5] J.G. Hacia, L.C. Brody, M.S. Chee, S.P. Fodor, F.S. Collins, Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis [see comments], *Nature Genet.* 14 (1996) 441–447.
- [6] D.J. Lockhart et al., Expression monitoring by hybridization to high-density oligonucleotide arrays, *Nature Biotechnol.* 14 (1996) 1675–1680.
- [7] M.S. Guyer, F.S. Collins, How is the Human Genome Project doing, and what have we learned so far?, *Proc. Natl. Acad. Sci. USA* 92 (1995) 10841–10848.

- [8] G.D. Schuler et al., A gene map of the human genome, *Science* 274 (1996) 540–546.
- [9] L. Rowen, G. Mahairas, L. Hood, Sequencing the human genome, *Science* 278 (1997) 605–607.
- [10] C. Fields, M.D. Adams, O. White, J.C. Venter, How many genes in the human genome?, *Nature Genet.* 7 (1994) 345–346.
- [11] F.S. Collins et al., New goals for the U.S. Human Genome Project: 1998–2003, *Science* 282 (1998) 682–689.
- [12] O. Ermolaeva et al., Data management and analysis for gene expression arrays, *Nature Genet.* 20 (1998) 19–23.
- [13] M.S. Boguski, G.D. Schuler, ESTablishing a human transcript map [news], *Nature Genet.* 10 (1995) 369–371.
- [14] G.D. Schuler, Pieces of the puzzle: expressed sequence tags and the catalog of human genes, *J. Mol. Med.* 75 (1997) 694–698.
- [15] Y. Chen, E.R. Dougherty, M.L. Bittner, Ratio-based decisions and the quantitative analysis of cDNA microarray images, *Biomed. Opt.* 2 (1997) 364–374.
- [16] J. Khan et al., Expression profiling in cancer using cDNA microarrays, *Electrophoresis* (1998) (in press).
- [17] J. DeRisi et al., Use of a cDNA microarray to analyse gene expression patterns in human cancer, *Nature Genet.* 14 (1996) 457–460.
- [18] N.L. Simone, R.F. Bonner, J.W. Gillespie, M.R. Emmert-Buck, L.A. Liotta, Laser-capture microdissection: opening the microscopic frontier to molecular analysis, *Trends Genet.* 14 (1998) 272–276.
- [19] J. Khan et al., Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays, *Cancer Res.* 58 (1998) 5009–5013.
- [20] R.F. Service, Will patent fights hold DNA chips hostage?, *Science* 282 (1998) 397.